Biweekly Report 3

Improving the dynamic comparison framework, expanding upon the ideas discussed in biweekly report 2 and theorising digital organisms

Raunak Narwal

Department of Mathematical Sciences

Indian Institute of Science Education and Research, Mohali, 130406, Punjab

October 18, 2025

General Modular Framework and how to run the files on local machine

We integrated the two files, computing topological similarity and dynamic comparison, into a single automated workflow, *runall.py*. This script takes two network files (in *.kgml* or *.csv*) as input and runs both persistence and dynamic analyses. For all the uses one can type python *runall.py* to see the help.

```
(venv) C:\Users\rishu_narwal\Desktop\MPI-CBG>python run_all.py
usage: run_all.py [-h] --graphA GRAPHA --graphB GRAPHB [--nameA NAMEA] [--nameB NAMEB] [--outdir OUTDIR] [--maxdim MAXDIM] [--p P]
                  [--convert-kgml]
run_all.py: error: the following arguments are required: --graphA, --graphB
```

To run the files on your local machine, clone this repository here, then run the following commands on your command prompt.

```
.venv\scripts\activate
pip install -r requirements.txt
python run_all.py --graphA --graphB --outdir
```

Improving the Dynamic Comparison framework

Our current dynamic comparison framework provides a simplistic functional baseline for simulating and comparing the behavior of two networks, it makes several simplifying assumptions that could limit its biological realism, current limitations include simplified reaction kinetics, all our reactions are currently happening with a rate constant of 1,

1

assuming uniform first order kinetics, this oversimplifies the system. Incomplete Stoichiometry representation, each reaction decreases substrate concentration and increase the product concentration by one unit, ignoring actual coefficients and multi substrate interactions. Another drawback of current workflow is that the framework aligns species solely by index order rather than by name or KEGG ID, which can produce misleading comparisons.

Some of the planned improvements inculde incorporating proper mass action and kinetics by parsing kinetic constants and rate laws from KGML or any external parameter file. Construct a full stoichiometric matrix that supports reactions of multiple substrates and products and that matches and compares only common species between networks based on KEGG IDs or names to ensure valid comparisons.

Pipeline Changes

The refactored script now has a clear and modular design to imrpove debugging and future expansions. Stoichiometry is now fully supported by the KGML stoichiometry attribute, encoded into a proper matrix and reaction rates now follow mass action kinetics, replacing the previous assumptions. Parameter handling is made optional, monte carlo runs for robustness analysis. Species matching is improved by aligning either KEGG IDs or names of species, preventing misaligned trajectory comparisons. CLI is improved to contain following flags

```
--params FILE
--tmax Max simulation time
--npoints Number of time points
--method ODE solver (RK45, BDF)
--steady_threshold Stop early if system reaches steady state
--mc Monte Carlo runs for averaging
--atol, --rtol solver tolerances
```

Results and Comparison with previous framework We compared map00400 and map00900 pathways with our updated dynamic comparison framework, the overall similarity metrics remained in the same range as previous, which suggests that underlying stoichiometry and reaction connectivity dominate the network dynamics more stongly than the kinetic form. The new framework shows a **DTW avergae of 6.2, MSE average of 2.14 and correlation average of 0.13, compared to 5.07 , 0.33 and 0.19 in the earlier version**. This increase in the averages reflects a higher dynamic distance, likely due to changes in the numerical solving percision and improved stoichiometric parsing.

Some compounds *(cpd:C00279, cpd:C00251, cpd:C00108)* maintained consistent patterns, with correlations above 0.9, which indicates that both frameworks capture the same trajectory trends despite numeric differences. *cpd:C00279 (ATP), cpd:C00944, cpd:C00493, and cpd:C03175* maintained strong positive correlations and low DTW/MSE values, which indicates that these species evolve in a similar temporal pattern across both the maps. These "similar acting" compunds likey corresponds to shared metabolic intermediates that play central role across pathways. This consistency of thheir trajectories suggests that the two network preserve common flux dynamics through these metabolites. Negative correlated species *(cpd:C00119, cpd:C00108, cpd:C00251)* indicate opposite flux trends between the networks, which tells pathway specific consumption or accumulation patterns. This often emerges because when some same metabolite serves distinct func-

tions in separate systems, as a substrate in one map but a product in another.

Comparaing graphs and charts
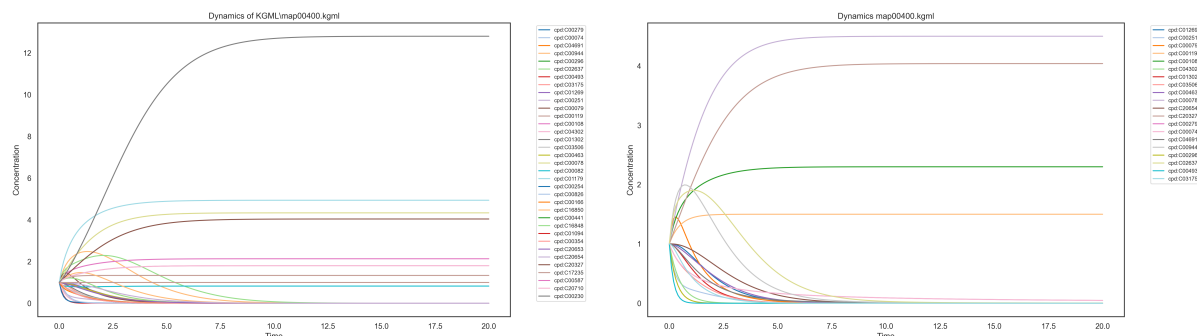
The improved framework only plots the common trajectories, this trajectory chart shows how concentrations of different compounds change over time. This is not comparing the two maps, it is the comparison of *map00400* trajectories plotted using two diffrent dynamic frameworks. Same for *map00900* is given below.
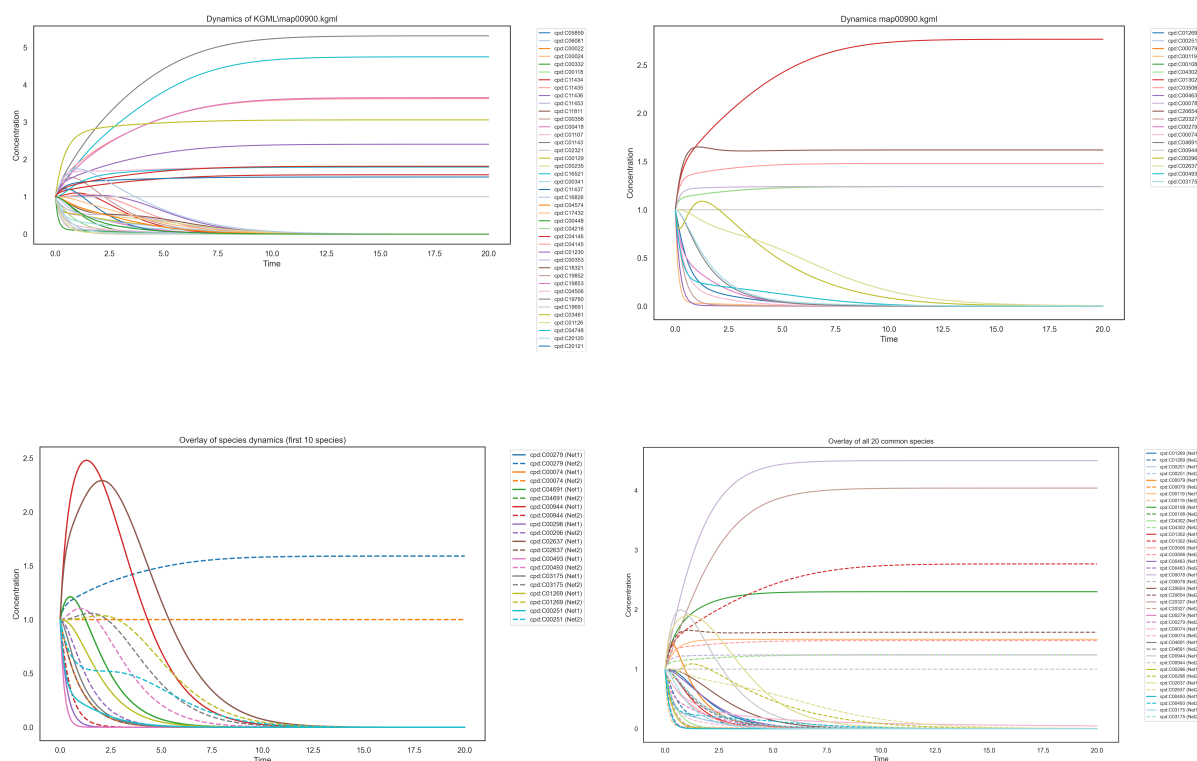
The Overlay of Trajectories seems very different, because the previous framework shows overlay of first 10 species index views, while new framework plots all the common species calculated by taking intersection of the KEGGIDs and then sorting the list out.
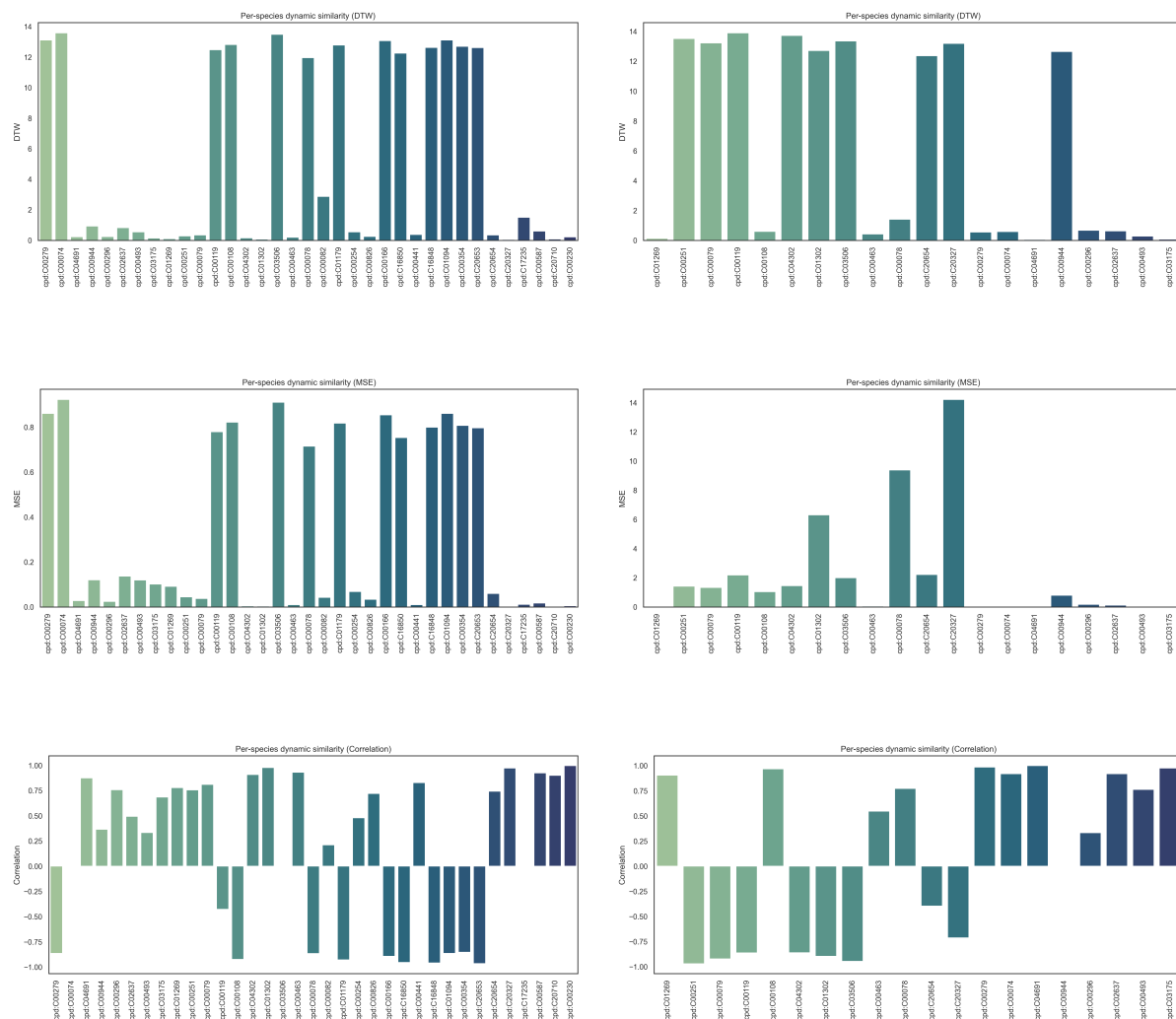
DTW measures the similarity between two temporal sequences which may vary in speed, MSE quantifies the average squared difference between corresponding points in two trajectories, while correlation assesses the strength and direction of a linear relationship between them. All three metrics together provide a comprehensive view of how similarly the species behave over time in the two networks. A low DTW score indicates that the two species follow similar shapes even if one is behind the other in reaching equilibrium. A lower RMSE means that the two species are comparable in concentrations over time and a correlation near 1 indicates both species rise and fall together suggesting similar regulation or flux direction.

Michaelis Menten Kinetics

Michaelis-Menten kinetics is a model describing how the rate of an enzyme-catalyzed reaction depends on the substrate concentration. It assumes a reversible reaction where an enzyme and a substrate form a temporary complex, which then turns into a product. The Michaelis-Menten kinetics is not included in the current implementation because of several reasons. It does not yet produce a measurable difference from the mass action model because reactions are using default k = 1 as rate constants are not provided with the .KGML files. Only a few XML files contain rate constants, seeing this we can think of some ways to calculate rate constants or take them up from some other website.

The saturation constant is given as $K_m = 1.0$. Without experimentally determined or literature-derived parameters, the Michaelis term

$$v = \frac{V_{max}[S]}{K_m + [S]}$$

reduces to near-linear behavior when $[S] \approx K_m$, mimicking the mass-action rate law. In real biochemical systems, V_{max} and K_m are enzyme-specific constants that define catalytic capacity. It is typically obtained from enzyme kinetics experiments, database values (for example BRENDA, SABIO-RK) or parameter estimation via model fitting against time series metabolomics data (measurement and analysis of all small-molecule metabolites produced by metabolism within biological samples). Thus k = 1 is a placeholder, it acts like a neutral scaling factor to maintain numeric stability, it preseves network topology, it allows for dynamic comparison without requiring unkown biochemical dat. The Monte Carlo runs are also producing the same results because intial conditions and parameters are all the same. MC becomes functionally meanigfull when perturbations like random scaling and kinectic constants are introduced to evaluate robustness and sensitivity of dynamic similarity metrics. **Estimating rate constants**
We can try to make our framework more biologically relevant by estimating rate constants using some approaches such as using computational parameter estimation methods, this might be time taking and maybe the trade off isn't fruitfull enough, so we will take estimating rate constant as an agenda point for future reports. One common approach is to fit the reaction model to avaible reaction data for metabolite concentrations over time by minimizing the error between the simulations and observed data points, using nonlinear least squares or global optimization. If we do no have direct experimental data for a map, then ballpark estimates may be taken from related literature. **Introducing animations for better visualisation**

Theorising digital organisms